DOCUMENT RESUME

ED 074 520 CS 200 454

AUTHOR	Zimmerman, Joel; And Others
TITLE	A Recognition Test of Vocabulary Using Signal-Detection Measures and Some Correlates of Word and Nonword Recognition.
INSTITUTION	Northwestern Univ., Evanston, Ill.
SPONS AGENCY	Office of Naval Research, Washington, D.C. Personnel and Training Research Programs Office.
PUB DATE	Feb 73
NOTE	58p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*College Students; *Educational Research; Educational Testing; Orthographic Symbols; *Vocabulary; *Word Frequency; *Word Recognition

ABSTRACT

In the first of three experiments, university undergraduates were presented a list of 300 words and 100 nonwords in two sessions. Their confidence that an item was a word was indicated for each item on a six-point scale. This experiment demonstrated the feasibility of creating a recognition test of vocabulary. In Experiment 2, 100 items were chosen to form a subtest, and the subtest was cross validated on a new sample of subjects. The test in Experiments 1 and 2 were scored using signal-detection measures. The primary criterion, SAT (verbal) scores, correlated approximately .60 with the test scores. In Experiment 3 subjects scaled the words and nonwords for four psychological attributes. These were submitted to a stepwise regression analysis with the confidence ratings from Experiment 1 as the dependent variable. It was concluded that associability, frequency, orthography, and pronounceability all may be components of word recognition. However, only frequency was found to be a significant predictor of the confidence of recognition of nonwords. (Author)

A RECOGNITION TEST OF VOCABULARY USING SIGNAL-DETECTION MEASURES AND SOME CORRELATES OF WORD AND NONWORD RECOGNITION

Joel Zimmerman, Paul K. Broder,
John J. Shaughnessy and Benton J. Underwood

Northwestern University

February 1973

Project NR 154-321
Sponsored by
Personnel & Training Research Programs
Psychological Sciences Division
Office of Naval Research
Arlington, Virginia
Contract No. N00014-67-A-0356-0010

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified

1. ORIGINATING ACTIVITY (Corporate author) Northwestern University Evanston, Illinois	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP ------------

3. REPORT TITLE
A RECOGNITION TEST OF VOCABULARY USING SIGNAL-DETECTION MEASURES, AND SOME CORRELATES OF WORD AND NONWORD RECOGNITION

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical report

5. AUTHOR(S) (First name, middle initial, last name)
Joel Zimmerman, Paul K. Broder, John J. Shaughnessy and Benton J. Underwood

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
February, 1973	51	16

8a. CONTRACT OR GRANT NO. N00014-67-A-0356-0010 b. PROJECT NO. NR No. 154-321 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ------------ 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) --------------

10. DISTRIBUTION STATEMENT
Approved for public release; distribution unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
---------------	Code 458 Personnel & Training Research Program, Office of [illegible] Research

13. ABSTRACT

In [illegible] of [illegible] s, [illegible] university undergraduates [illegible] presented a list of [illegible] words and 10[illegible] nonwords in two [illegible]. Their confidence that an item was [illegible] was indicated for each item on a 6-point scale. This experiment demonstrated the feasibility of creating a recognition test of vocabulary. In Experiment II, 100 items were chosen to form a subtest, and the subtest was cross validated on a new sample of subjects. The tests in Experiments I and II were scored using signal-detection measures. The primary criterion, SAT (verbal) scores, correlated approximately .60 with the test scores. In Experiment III subjects scaled the words and nonwords for 4 psychological attributes. These were submitted to a stepwise regression analysis with the confidence ratings from Experiment I as the dependent variable. It was concluded that associability, frequency, orthography, and prounounceability all may be components of word recognition. However, only frequency was found to be a significant predictor of the confidence of recognition of nonwords.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6801

Unclassified
Security Classification

Unclassified
Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Memory Recognition Frequency information						

Unclassified
Security Classification

A RECOGNITION TEST OF VOCABULARY USING SIGNAL-DETECTION MEASURES, AND SOME CORRELATES OF WORD AND NONWORD RECOGNITION

Joel Zimmerman, Paul K. Broder,

John J. Shaughnessy and Benton J. Underwood

Northwestern University

February 1973

Project NR 154-321
Sponsored by
Personnel & Training Research Programs
Psychological Sciences Division
Office of Naval Research
Arlington, Virginia
Contract No. N00014-67-A-0356-0010

A RECOGNITION TEST OF VOCABULARY USING SIGNAL-DETECTION MEASURES, AND SOME CORRELATES OF WORD AND NONWORD RECOGNITION

Joel Zimmerman, Paul K. Broder

John J. Shaughnessy, and Benton J. Underwood

Northwestern University

Abstract

In the first of 3 experiments, university undergraduates were presented a list of 300 words and 10[illegible] nonwords in two sessions. Their confidence that an item was a word was indicated for each item on a 6-point scale. This experiment demonstrated the feasibility of creating a recognition test of vocabulary. In Experiment II, 100 items were chosen to form a subtest, and the subtest was cross validated on a new sample of subjects. The tests in Experiments I and II were [illegible] using signal-detection measures. The primary criterion, SAT (verbal) scores, correlated approximately .60 with the test scores. In Experiment III subjects scaled the words and nonwords for 4 psychological attributes. These were submitted to a stepwise regression analysis with the confidence ratings from Experiment I as the dependent variable. It was concluded that associability, frequency, orthography, and pronounceability all may be components of word recognition. However, only frequency was found to be a significant predictor of the confidence of recognition of nonwords.

A RECOGNITION TEST OF VOCABULARY USING SIGNAL-DETECTION MEASURES, AND SOME CORRELATES OF WORD AND NONWORD RECOGNITION[1]

Joel Zimmerman, Paul K. Broder

John J. Shaughnessy, and Benton J. Underwood

Northwestern University

In some recent experiments (e.g., Carroll, 1971) subjects have been presented a set of words and asked to judge the frequencies with which such words occur in printed English. This procedure rests on the assumption that subjects keep some kind of "tally" in their day to day interaction with printed English. It also assumes, however, that subjects have truly seen the words which are being tested, an assumption which may have no basis in truth for very low frequency words.

Assume there was a sample of words which were distributed along a continuum according to their true frequencies of occurrence in printed English, from very frequent to very infrequent. For any hypothetical subject, this sample of words could be thought of as a monotonically decreasing scale of word familiarity. Words at the beginning of the continuum, would surely be recognized. There would be some point on this continuum, though, such that all words past such a point would never before have been seen in printed English by this subject. In judging frequency, the subject ought to rate each word past this point a "zero," and we would say at this point that the subject now fails to recognize any further words.

The concern of the present study is not with the ability to judge word frequency, but with the matter of recognition. At least two lines of

[1]The authors wish to thank Dr. Robert Sekuler for his advice regarding signal detection theory.

investigation become apparent. The first of these concerns individual differences in recognition, and the second concerns the psycholo[illegible]cal processes involved in recognition.

Considering, again, the continuum of words with differing background frequencies, it would be reasonable to suppose that the point on this continuum at which a subject would no longer recognize words would be quite different from subject to subject. One might suppose that as a subject's vocabulary increased, he would be able to proceed farth[illegible]nd farther dow[illegible] t[illegible] c[illegible] before [illegible]aching the hypothetical point. This is to say that subjects with better vocabularies should be able to recognize more words, an intuitively appealing statement which would probably be accepted by most people without any prefatory rationale.

One purpose of this experiment was to determine whether a test might be developed to assess vocabulary skills based on this principle, i.e., that people with better vocabularies will recognize more words as words. Subjects were thus shown a series of words and asked whether or not they recognized each one. To make the task realistic, a fourth of the items presented were nonwords, which the subject was not expected to recognize as words. The subject expressed his confidence in each recognition judgment through the use of a category rating scale. Word recognition ability was assessed using a measure derived from signal detection theory. This measure served as an estimate of the subject's ability which theoretically would be free from the effect of a subject's [illegible]s in using the rating scale.

The second line of interest in the experiment was to inquire into the psychology of word recognition. On what basis does a subject decide that he recognizes a word? Further consideration will be given to this problem in the introduction to Experiment III.

Experiment I

Method

Materials. The complete test form was composed of 400 stimuli, of which 300 were words and 100 were nonwords. A representative sample of 400 words was drawn from a standard English language dictionary (G. & M. Merriam Co., 1963). No compound words, hyphenated words, or homographs were allowed. Words had to be at least four, and not more than 10 letters in length. From this pool, 100 words were selected randomly and clustered by number of syllables. Corresponding syllables of words within these clusters were then interchanged in a random manner to produce nonwords. One-syllable words were arbitrarily divided into two parts and these parts were interchanged. Nonwords which resulted in combinations of syllables which, in the opinion of the first author, were extremely difficult to pronounce or which resulted in true words were subjected to a second or third random interchange of syllables with other such items. After three such interchanges, about five nonwords were still extremely difficult to pronounce, and minimal changes were made in the letters of these items to make them pronounceable. By this method, 100 nonwords were created which had about the same average length, number of syllables, and letter frequency as did the 100 real words from which they came.

An unabridged dictionary (Stein & Urdang, 1967) was then checked to affirm that these items were not, in fact, words.

The 300 remaining words and the 100 nonwords were randomly placed into 10 groups of 40 with no restrictions. The 40 items in each group appeared in lower-case type in two columns of 20 items each on plain 8 1/2 X 11 in. paper. Placement of an item within a column was random. At the top of each page was space for the subject's name and the date. There was also an explanation of the six-point scale which the subject was to use as follows:

1 means you are very sure this <u>is not</u> a word

2 means you think this is not a word

3 means you guess this is probably not a word

4 means you guess this is probably a word

5 means you think this is a word

6 means you are very sure this is a word

Following each word on the page were the numbers from 1 to 6 in a row.

A test booklet contained each of the 10 pages of 40 items. The pages were arranged according to a 10 X 10 Latin square to assure that each page would be viewed in each position an equal number of times across a group of 10 subjects. Since subjects were to judge only five pages per day, each group of 10 test booklets constructed from a given Latin square was duplicated except that the first five pages and the last five pages were interchanged. Ten Latin squares were selected randomly, allowing for the construction of 200 test booklets.

Procedure. Subjects were 200 Northwestern University undergraduates who served in this experiment either voluntarily or to fulfill a course requirement. Each subject served on two successive days, judging 200 items in five pages of a test booklet during each session. Instructions were read which stated that this was an attempt to gather information on word familiarity prior to using these items in constructing a vocabulary test. Subjects were told that some of the items were not really words, but that most were real words. The rating scale was explained, and the subjects were instructed to read through the items in order, circling one of the six numbers next to each item to indicate their certainty that a given item was or was not a word. The instructions made it clear that a word was not to be doubted on the basis of spelling, i.e., all items were to be considered as correctly spelled. Subjects were asked to indicate their high school rank at graduation, and their scores on the verbal section of the Scholatic Aptitude Test.

Subjects were tested in groups varying in size from 1 to about 25, and at times convenient to them. The second session for a subject occurred from 18 hours to 30 hours after the first. For most subjects, the period was 24 hours. One subject failed to return within these time bounds, his first day's data were discarded, and an identical test booklet was constructed and given to the next subject. During analysis, two subjects were found to have skipped a page, and their data were discarded and replaced. One subject was found to have clearly and consistently reversed the six-point scale. This subject's data were corrected and retained.

One word, "sabadilla", was misspelled and printed as "sabadilia" on the test sheets. The word was scored as though it had been printed correctly, and scores for this item were retained in all analyses.

Results

Scaling results. The 100 nonwords and the 300 words are listed alphabetically in the Appendix. Following each item is the mean judgment given the item and the standard deviation of the judgments. Also included in this table is a measure of internal consistency which is the correlation between the judgment made on this item, and the mean judgment made on all such items (nonwords or words) by each subject.

The mean of all judgments given to nonwords was 2.92 and the standard deviation was 1.30. The mean of all judgments given to words was 4.90 with a standard deviation of 1.53. Thus, there was slightly more variability overall in the judgments made on words than on nonwords. If the measure taken is the mean rating given to nonwords and words by each subject, however, this conclusion is modified. The mean ratings given to 100 nonwords by subjects varied from 1.16 to 4.75, a range of 3.59, and the standard deviation of these means was .70. The mean ratings given to words varied from 4.18 to 5.78 over subjects, a range of only 1.60, and the standard deviation of these means was only .32. The reason for these results seems to be that each subject was more consistent in his judgment of nonwords than in his judgment of words, but that different subjects were more variable in choosing a portion of the scale within which they chose to rate nonwords.

Prediction of criteria. [illegible] the basis of their availability, two criteria were chosen to [illegible] alidation of this test, h[illegible] school rank at graduation (HSR) and the verbal score of the Scholastic Aptitude Test (SAT). HSR was scored as the percent of the class not ranking as high as the subject. SAT scores are nationally normed scores ranging from 200 to 800. We assumed that the HSR scores would represent a measure of some sort of general ability, while the SAT scores would represent some measure of ability more specifically verbal in nature. SAT scores were, therefore, the more important of the two criteria. For this reason, only prediction of SAT scores will be discussed at length in this report. Some of the correlations with HSR are listed, however, in Table 1.

Subjects were requested to report their HSR and SAT scores at the time of the test. Eight subjects failed to report each of the scores, and to facilitate analysis, these subjects were assigned the nearest integer to the mean of the scores of the 192 subjects from whom scores had been obtained. This resulted in a mean reported SAT score of 610.40 with a standard deviation of 79.45 over the 200 subjects. HSR scores had a mean of 89.16 and a standard deviation of 10.94.

On the most unassuming level of analysis, one might suppose that those subjects who had the highest vocabulary skills would be those who most confidently recognized words as words, and nonwords as nonwords. This would lead one to expect a positive correlation between the criterion and the mean rating given to words, and a negative correlation between the criterion and the mean rating given to nonwords. The correlations between SAT scores and the mean judgments of words and nonwords were .20

and -.13, respectively. Only the prediction of SAT scores from word rating means was significantly different from zero (t=3.07, df=198, p<.01).

Such an attempt at prediction was, of course, naive. For one, it assumed that subjects with higher vocabulary skills would give both higher mean ratings to words and lower mean ratings to nonwords. This would predict a negative correlation between mean ratings given to words and nonwords. The computed correlation between these measures, however, was .76. At least part of this correlation must reflect the subjects' biases in using high or low numbers on the scale independent of the nature of the particular stimulus being judged.

A measure which could overcome some of this bias would be the difference between the mean judgment for nonwords and the mean judgment for words for each subject. With increased confidence in both word and nonword recognition, the difference between the word and nonword judgment means for a subject should have increased, and ideally this should have been independent of the subject's bias in using some part of the scale. In fact, using this measure, prediction of SAT scores increased slightly to .31.

This measure is still deficient, however. It does not take into account the variability with which a subject made judgments. A given difference between word and nonword means increases in significance as the variability around those means decreases. Therefore, a way to increase the usefulness of the difference measure should be to standardize it with respect to the subject's variability in making judgments. To do this, the difference between word and nonword mean judgments for each subject was divided by the square root of the pooled variance of that subject's judgments.

around those means. When this was done, the correlation between this measure (which will hereafter be referred to as d_s, for standardized difference) and SAT scores was found to be .48. This increase in prediction over .31, obtained by accounting for the subjects' variabilities in judgments, was statistically significant (t=2.23, df=197, p<.05).

Signal detection theory. Signal detection theory was first applied to verbal materials by Egan (1958), and is probably most thoroughly explicated by Green and Swets (1966). Reviews of the use of signal detection theory in memory experiments have been provided by Banks (1970) and Lockhart and Murdock (1970). The purpose of a signal-detection analysis is to separate two components of a subject's behavior, his sensitivity in responding to a stimulus, and his bias in responding. As indicated by Banks, bias in responding in a verbal recognition task is often related to the idea of guessing, and analysis by signal-detection measures becomes, at very least, a sophisticated way of correcting for guessing. The application of signal detection theory to the present experiment is relatively straight forward. It is assumed that for the population of words and nonwords, the degree of a subject's confidence of recognition would be normally distributed around some mean for words and some mean for nonwords. The distributions are assumed to be normal and of equal variances.

As a test of whether the present data have met the assumptions of signal detection theory, a memory operating characteristic (MOC) curve has been plotted for the group as a whole, and is presented as the top line in Figure 1 (ignore, for now, the other data in the figure). The hit rates and false alarm rates have been transformed to standard unit normal distribution equivalents. According to Lockhart et al., a straight line in a MOC curve is commonly taken as evidence that the assumption of normal distributions has been met, but since large deviations from normality will p' Mc :ves which appear to be straight lines, this is not a criticai test. Nonetheless, the straight line best fit (by the method of least squares) for these data is quite good, $\underline{r}$=.998. A line with a slope of 1.0 is taken to be evidence for the assumption of equal variances. It has already been stated, however, that the variances for nonword and word judgments were not equal, and this is reflected by the fact that the line has a slope of .55.

The value commonly suggested as a nonparametric measure of the subject's recognition sensitivity is the area under the subject's MOC curve (A). This value is used as an index of d', which corresponds to the separation of the recognition confidence distributions for words and nonwords. It is just this separation which we were attempting to measure by the use of d_s described above. MOC curves were derived and A was computed for every subject. The contention that A and d_s are theoretically equivalent was supported by the high correlation between these measures, $\underline{r}$=.96. The correlation between SAT scores and A was .44, about the same as prediction of SAT scores by d_s ($\underline{r}$=.48).

It is a theoretical question whether any single measure can be derived from confidence rating data to represent the subject's bias

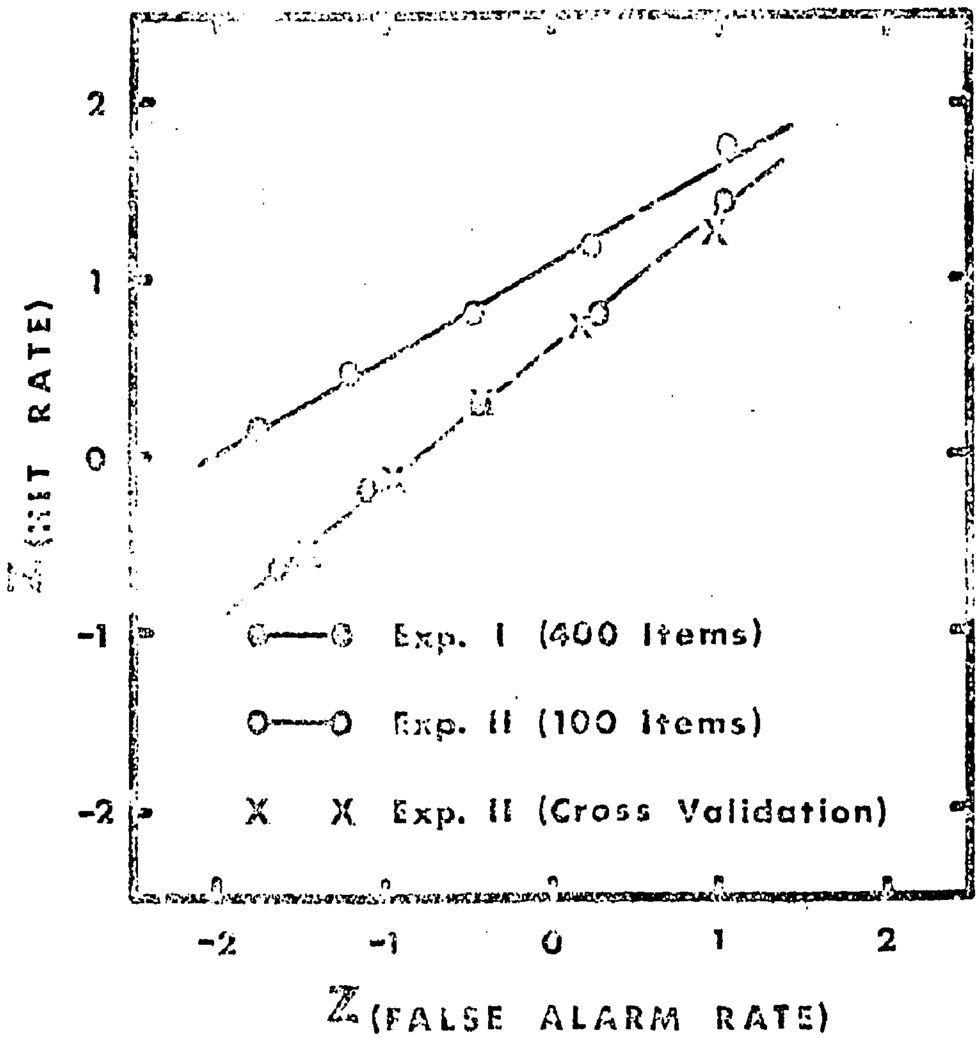

Fig. 1. Memory operating characteristic curves for the group of 200 subjects' ratings on 400 items, their ratings on the 100 item subtest, and for the independent group of 42 subjects' ratings on the 100 item subtest.

in responding (commonly referred to as β or B). One measure used by McNicol and Ryder (1971), however, indexes B as that unique point on the subject's MOC curve (determined by linear interpolation) at which the hit rate and false alarm rate sum to 1.0. In a situation perceived by the subject to have equal a priori frequencies of words and nonwords, and equal "payoffs" for hits and correct rejections, this point could be conceptualized as a theoretical point on the confidence rating scale. This point would represent the rating for an item for which there was maximal uncertainty as to whether it was a word or a nonword. This value, B, was calculated from each subject's MOC curve.

Rather than working from MOC curves, we have preferred to index subject bias in a more direct way. Under an assumption of normal distributions of equal variance, the B measure described above would be identical to the point on the rating scale midway between the means of the word and nonword confidence distributions. This value (henceforth to be referred to as M, for midpoint) was also computed for every subject by taking the simple average of the mean rating for nonwords and the mean rating for words. The two measures, B and M, were found to correlate highly, $r=.95$. If all the assumptions of signal detection theory had been met, these measures of subject bias would ideally be independent of the sensitivity measures (A and d_s) and would not predict the criterion scores (SAT). In fact, B correlated -.26 and -.29 with A and d_s, respectively, and M correlated -.18 and -.21 with A and d_s, respectively. These correlations were all significantly different from zero ($p<.05$). The correlation between B and SAT scores was -.07 ($p>.05$), and between M and SAT scores was -.03. Thus, it may be concluded that the bias measures, B and M, were not predictors of SAT scores, but they were not entirely independent of the sensitivity measures, A and d_s.

There are three important points to be gleaned from these d- First, the d_s measure is theoretically and functionally equivalent to the sensitivity measure postulated within the theory of signal detection. Secondly, the sensitivity measures have been shown to have low correlations (r=-.23 on the average) with the measures B and M. Insofar as B and M index subjects' biases it can be concluded that, while the sensitivity measures were not completely independent of subject bias, they are, at least, only slightly affected by bias. Finally, the bias measures did not significantly correlate with the criterion scores (r=-.05 on the average). Therefore, the significant prediction of criterion scores by the sensitivity measures was in no way due to, and, in fact, must have been in spite of, the degree to which the sensitivity measures reflected subject bias in the use of the scale.

Official SAT scores. Subsequent to the above analyses, the question was raised as to the veracity of the subject's reports of their criterion scores. We were able to obtain official school records of 156 SAT and 155 HSR scores. The subject's reported SAT scores, and those provided by the school for these 156 subjects correlated .90. It is interesting to note that the scores reported by these 156 subjects were just slightly lower on the average than the scores of the other 44 subjects, yet averaged almost 19 points higher than the scores obtained from the university records. The correlation between the d_s and A measures and the official

SAT scores were slightly higher than those with the full sample of reported scores, .56 and .54, respectively.

Reported and official HSRs were about equal in magnitude, and correlated .38 over the 155 subjects for whom the scores were available. An average was taken of the correlations between reported HSR scores and the two measures d_s and *A*. This average correlation was also calculated using official HSR scores, using reported SAT scores, and using official SAT scores, and all these average correlations are summarized in the first column of Table 1.

Testing effects. In laboratory tests of recognition, changes in performance as a function of the test interval or the testing procedure have been of considerable interest, and have been referred to as testing effects (e.g., Underwood & Freund, 1970; Underwood, 1972). In the present experiment, the judgments of the items on each page were systematically balanced over positions within and between days. Thus, changes in subject judgments over pages and days could be examined free of any confounding by specific item groups. The mean word judgment given by all subjects on each of the 10 pages of the test booklet ranged from 4.86 to 4.93 in no systematic or statistically significant way. In the judgment of nonwords, however, the page effect (F=2.57, df=4,796) and the day effect (F=5.07, df=1,199) were significant (p<.05) though their interaction was not (F=1.70, df=4,796, p>.05). The mean judgments of nonwords on the two days of testing were 2.89 and 2.95. Across the five pages of a test

Table 1

Average Prediction of Criterion Scores by the Sensitivity Measures, $\underline{d}_s$, and $\underline{A}$, for the 400-Item Test, the 100-Item Subtest, and the Cross Validation of the 100-Item Subtest

	400 Items	100 Item Subtest	Cross Validation
SAT - reported	.46 (200)	.66 (200)	.64 (36)
SAT - official	.55 (156)	.69 (156)	.58 (18)
HSR - reported	.26 (200)	.32 (200)	.19 (40)
HSR - official	.25 (155)	.31 (155)	.08 (17)

Note: Number of cases is indicated in parenthesis booklet (averaged over the two days) the mean judgment given to nonwords rose slowly but systematically from 2.88 to 2.96. Though statistically reliable, the magnitude of these effects is so small as to seem empirically unimportant. The measure *M*, being a midpoint between the word and nonword means also had to increase slightly over days and pages as a result of the increase in nonword means and the relative stability of word means. If *M* were considered to be a measure of subject bias, the small testing effects in *M* could be interpreted as a change in subjects' biases over days and pages resulting from a slightly but reliably increasing tendency to guess that a nonword might be a word.

The d_s measure was calculated for every page of each subject's test booklet. The average d_s varied from 1.44 to 1.53 in no systematic or statistically significant way. Thus, there was no evidence for any meaningful change in sensitivity over the testing interval.

To these data can be added the correlations of the subjects' nonword means, word means, and d_ss calculated separately for day 1 and day 2. These correlations were .84, .71, and .58, respectively, which though high, are not particularly noteworthy for reliabilities. In summary, it is concluded that subject performance on this test was relatively stable, and whatever systematic changes in behavior did occur over the testing interval were small, indeed.

Dichotomous judgments. In many recognition studies (e.g. Underwood, 1972) the subject's task has been to respond with a "yes" or "no," and items have been scored as either right or wrong. It is not unreasonable to ask whether this more simple way of responding and scoring would have produced results comparable to or better than those which were achieved on this test using 1 to 6 confidence ratings. In accordance with the labelling of the points on the rating scale, it was assumed that any item which had been rated 1, 2, or 3, would have been classified as a nonword if the subject had been making simple dichotomous judgments. Likewise, items rated 4, 5, or 6 presumably would have been classified as words in such a task. All of the data for all subjects were accordingly transformed into dichotomous judgments and scored as correct or incorrect. Several procedures were used to evaluate these data, including simple raw scores, correction for guessing scores, and signal-detection scores. These scores uniformly resulted in skewed distributions, and relatively low correlations with SAT scores. It was concluded that this was not a valuable way to proceed, and insofar as this procedure truly mimicked what would be obtained in a "yes"-"no" test on nonwords and words, such dichotomous responding does not provide data equal in quality to that obtained by the confidence-rating method.

Experiment II

The first experiment demonstrated the practicability of constructing a vocabulary test based on absolute judgments of word recognition, and scoring it in accordance with the methods of signal detection theory. The purpose of the second experiment was threefold. First, the test was to be decreased in length so that a subject might easily be given the instrument in one session. Secondly, "bad" items were to be eliminated so as to increase the overall predictive power. In the remainder of the paper, the collection of items retained after decreasing the test in length will be referred to as the subtest. This subtest was initially evaluated by deriving scores for the subjects in Experiment I as though these were the only items which had been judged. Third and finally, the subtest items were to be assembled as a separate test and judged by an independent group of subjects. This will be referred to as the cross validation.

Method

Subtest. From the pool of 400 items, 26 nonwords and 74 words were chosen to be used as a 100-item subtest. To evaluate this subtest, subject protocols from Experiment I were rescored as though the subjects had rated only these 26 nonwords and 74 words.

Three criteria were used to select the subtest items. The primary consideration was to obtain a set of items such that frequency distributions of the mean judgments on the items (from Experiment I) would be approximately normally distributed for nonwords and for words. Accordingly, the rating scale was divided into units of 0.5 width. From the eight intervals along the scale beginning with 2.00 to 2.49 and ending with 5.50 to 5.99, the following numbers of words were selected for the subtest: 1,6,12,18,18, 12,6, and 1. From the six intervals beginning with 1.50 to 1.99 and ending with 4.00 to 4.49 the following numbers of nonwords were selected: 1,3,9,9,3, and 1. The second selection criterion concerned an item's discriminability. The rating on an item was correlated with reported SAT scores over all 200 subjects. Within a rating-scale interval, those items were chosen which best discriminated among subjects according to this index. The third criterion for selection was a high internal consistency index, as described in the Results section of Experiment I.

Cross validation. In order to cross validate the subtest, the 100 items were assembled separately into a test booklet. The items were placed randomly onto five pages, 20 items per page. The 20 randomly ordered items appeared in lower-case type in a single column. Except for the lesser number of items, the test sheets were exactly as described in Experiment I. The five test sheets were presented in the same order to every subject. The position (from 1 to 100) of each item which appeared in the subtest booklet is indicated in the last column of the Appendix. Items 1 through 20 appeared on page 1, 21 through 40 on page 2, and so on.

The cross validation sample consisted of 42 subjects who took the subtest as partial fulfillment of a course requirement at Northwestern University. Subjects were tested in groups ranging in size from 1 to about 25. They were told that this experiment was being done to develop a new kind of vocabulary test, and they were given instructions on rating the items as were subjects in Experiment I. Subjects were asked to report their SAT and HSR scores. No subjects were lost or replaced for any reason.

Results

Subtest. The most obvious consequence of selecting items for the subtest was to remove words with very high confidence ratings. Accordingly, the mean word rating was changed from 4.90 in the full 400-item form to 4.01 in the 100-item subtest. The standard deviation was increased slightly from 1.53 in the original to 1.62 in the subtest. The change in the mean judgment and standard deviation for nonwords was slight, from 2.92 to 2.99 and from 1.30 to 1.35, respectively. Considering the mean word and nonword judgment for each subject, the result was the same. The standard deviation for the subjects' mean word judgments changed from .32 to .59, whereas the standard deviation of subjects' mean nonword judgments changed from [illegible] to .73.

Some other observations may be made to indicate the degree to which the subtest was representative of the complete test. The subject's scores for the full test of 400 items and their scores on the 100-item subtest were correlated for six measures. The correlations for the mean rating given nonwords, the mean rating given words, and the bias measures, M and B were .95, .94, .98, and .96, respectively. For these four measures, then, the subtest was highly representative of the full test. The corresponding correlation for $\underline{d}_s$ was .77 and for A was .76. The correlations on these sensitivity measures were obviously not as high as those for the other measures, and this is evidence that the subtest was primarily functioning to change slightly and differentially the estimates of the subjects' abilities.

The group MOC curve based on the subtest is shown as the bottom line in Figure 1. The movement of the line toward the major diagonal indicated that the average sensitivity as measured by the 100-item subtest was lower than that measured by the full test of 400 items. This was the result of removing the easy words which had served to indiscriminately raise all subjects' scores. In fact, the average As for the full test and subtest were .83 and .69 in that order, and the average $\underline{d}_s$s were 1.42 and .73. The increase in the slope of the line to .76 indicated that there was less difference between the word judgment variability and nonword judgment variability in the 100-item subtest than had been present when all 400 items were considered, and this is in accordance with the standard deviations of these judgments reported above. Again, the linear fit is quite good (r=.999), though this is not a particularly critical test of the normality assumption.

The sensitivity measures, d_s and A, correlated .98 and the bias measures, B and M, correlated .97 on the subtest, comparable to these same figures from the overall test (.96 and .95). The average inter-correlation of the sensitivity measures with the bias measures had been significantly greater than zero in the full test (r=-.23), but this correlation, r=-.05, did not differ significantly from zero in the subtest (p>.05). Referring again to Table 1, it can be seen that the correlation of reported and official SAT scores with the sensitivity measures increased considerably on the subtest, from .46 to .66 (N=200) and from .55 to .69 (N=156) in that order. The average correlation of reported SATs with the bias measures (B and M) had been -.05 on the full test, and was -.01 on the subtest. The corresponding values with regard to the official SAT scores were -.10 and -.05.

Cross validation. With the independent group of 42 subjects taking the subtest, the mean nonword rating, 2.98, and standard deviation, 1.46, were very comparable to those values calculated for the original sample of subjects (2.99 and 1.35, respectively). These values for the word ratings were 4.04 and 1.71, also comparable to those values obtained on the subtest with the first sample (4.01 and 1.62). The standard deviations of the mean judgments made by a subject for nonwords, .55, and for words, .44, were lower in this sample than in the original sample (.73 and .59, respectively). For the cross validation sample the split half reliability calculated over subjects for mean nonword judgments was .86, and for mean word judgments was .71. (Day by day reliabilities calculated for the 400 items in Experiment I had been .84 and .71 for nonword means and word means.) Additionally, the correlation between the mean rating on an item from the first sample and the mean rating for the same item from the cross validation was .90 for the 26 nonwords and .91 for the 74 words, and these may be taken as estimates of item reliability. It is concluded from these data that performance on the subtest was highly comparable for the original and cross validation samples of subjects.

The comparability of use of the scale by the two groups with respect to signal detection theory may be assessed with regard to the group MOC curve. The "X" marks in Figure 1 represent the MOC points for the cross validation group. The line of best fit to these points has not been drawn in, since it would be indistinguishable from that for the original sample, and for all practical purposes, it is apparent that the line from the original sample serves to describe the cross validation sample as well.

Out of 42 subjects, 36 reported SAT scores, and the university provided official scores for 18 of these. The mean of the 36 reported scores was 607.36, with a standard deviation of 73.34. Official SAT scores averaged over 33 points lower than the scores which were reported by these 18 subjects. The reported HSRs averaged 83.27 and had a standard deviation of 15.69. Reported and official scores correlated only .55 for SAT and .80 for HSR.

The two sensitivity measures, $\underline{A}$ and $\underline{d}_s$, correlated .99 in the cross validation group, but the .89 correlation between $\underline{B}$ and $\underline{M}$ was low in contrast to what we had come to expect. Table 1 summarizes the prediction of the criteria for this group. The average prediction of the reported SAT scores by the sensitivity measures was .64 and compared favorably with the .66 prediction in the first sample. The prediction of the official scores was disappointingly lower, having fallen from .69 to .58, though this difference was not significant ($\underline{t}$=.69, $\underline{df}$=168, $\underline{p}$>.05). One is tempted to attribute this fall to the small size of the cross validation sample for which official scores could be obtained (18). The average intercorrelation of the sensitivity and bias measures was -.18 ($\underline{p}$>.05) and the reported and official SAT scores were predicted by the bias measures with average correlations of -.02 and -.07, respectively.

Discussion

From the results of this experiment, we conclude that a recognition test of vocabulary scored through the use of measures derived from signal detection theory provides distinct promise as a tool for evaluating vocabulary skills. This test of 100 items is easily administered to the average student in about 15 minutes. The test is easier to take than the typical vocabulary test of the same length which usually involves reading a word and searching through several alternative definitions for the one which best fits. While correction for guessing procedures are still a matter of theoretical debate with regard to the usual multiple-choice format, this testing procedure yields a separate measure of bias (or guessing tendency) as well as a sensitivity measure. The resultant sensitivity measure correlates very acceptably (about .60) with scores from the verbal sections of the Scholastic Aptitude Test. It might be argued, in fact, that this correlation underestimates the validity of the test. The present test purports to be a measure of vocabulary, while the SAT is presumably a measure of verbal ability in a more general sense. Certainly, vocabulary skill must contribute in large part to the score on the SAT and no other convenient, dependable measure of vocabulary skill was available for use as a criterion. If this test were to be validated against some more direct measure of vocabulary ability, though, the estimate of its validity might be even higher.

Further, this is not meant to be the final version of a test. The primary purpose of this experiment was to demonstrate the feasibility of such an approach, not to provide a highly developed product. The items in the present test have been put through only one selection process, and improvements in the item pool could certainly be made. In line with signal detection theory, for example, it might be suggested that a pool of words and nonwords be developed which yield more equal variabilities in judgments. This should increase the validity of the $\underline{d}_s$ measure of sensitivity. Further work might also be done to improve the nature of the scale. The intervals in the rating scale have been assumed to be equal, but may be psychologically very different. A scaling of the intervals on the six-point scale into their proper psychological equivalents could reveal a transformation of either the judgment data, or the scale itself, which should serve to increase the validity of the measurement.

This paper also lends another level of generality to signal detection theory, a theory which is encountering widespread success in application to recognition situations of many kinds. To the best of our knowledge, the present technique for calculating d_s and M are not commonly in use. With two different samples of subjects, however, these measures were shown to correlate highly with A and B, respectively, as they have previously been calculated from subject MOC curves (see Green & Swets, 1966; McNicol & Ryder, 1971). Of particular relevance is the fact that d_s and M do not need to be calculated from MOC curves, which simplifies the computations of these measures considerably.

Experiment III

Up to this point, consideration has been given only to differentiating among subjects on the basis of how well they can recognize words and nonwords. No mention has been made of how such a process of recognition might be occurring. Yet, a knowledge of the psychological processes involved in differentiating between meaningful and nonmeaningful verbal stimuli would be of considerable importance.

Several investigators have proposed mechanisms to explain how words are recognized. McNulty (e.g., 1966) has proposed that recognition is accomplished through the learning of partial information. After exposure to a stimulus, the subject can not reproduce the stimulus in its entirety, but has retained some information about it. At the time of recognition, the subject generates the partial information and checks it for a match against the stimulus provided. If the partial information which the subject can generate is entirely matched by the stimulus, the subject accepts the stimulus and says he recognizes it. McNulty proposes that such partial information can be structural, such as individual letters, or associative, such as knowledge that an item was a member of some category. Since nonwords in

Experiment I were created from fragments of real words, their structures were sound. This point of view probably would hold, then, that subjects distinguished between words and nonwords on the basis of associative processes, or more properly, the lack of them. That is, perhaps the subject looked at a nonword, decided that there was nothing with which he could consistently associate it, and thereby gave it a low rating. Words, on the other hand, might have brought to mind familiar associations and were accordingly accepted as words. If this is a fair representation of McNulty's position, it might seem to predict that nonwords which produce consistent associations might be those which are most often mistaken for words.

Underwood and Freund (~~e.g.~~, 1968) ~~has~~ have proposed that frequency is the attribute which mediates recognition in a laboratory situation. Very simply, each time a stimulus is perceived in a study list it accumulates an additional frequency input. When an item is presented on a recognition test, the subject merely checks the frequency count on the item. If it is greater than zero, the item is recognized. A generalization of this theoretical position with respect to words would predict that the more often an item has been seen or heard, the more likely it is to be recognized as a word. This received some support from data in Experiment I which revealed that for 162 words, the mean confidence rating given a word correlated .30 with the Thorndike-Lorge "G" frequency. Nonwords, however, all should have frequencies of zero, and the only way that nonwords could be differentially recognized according to a frequency theory would probably be through some consideration of relative frequencies of combinations of letters or syllables making up the words.

Another consideration of word recognition comes from Smith and Haviland (1972). These investigators studied the question of why perception of words through brief tachistoscopic exposure was more accurate than perception of nonwords. They concluded that the perceptual unit of analysis of a word is the pronounceable English segment. For a nonword, however, the perceptual unit of analysis is the individual letter. This might suggest that subjects pronounce the items, and decide to identify an item as a word or nonword on the basis of pronounceability. Though all the words in Experiment I were purposely made at least moderately pronounceable, it might still be expected that as pronounceability of the items increased, subjects would be more likely to perceive them as words. Given that the items were not easily pronounced, the Smith and Haviland view might lead to the expectation that the subject then examined the individual letters, noting distinctions in what has been called the orthography of the word (Zechmeister, 1969).

Perhaps the most thorough consideration of the process of recognition of words and nonwords comes from a series of studies by Rubenstein and his coworkers (Rubenstein, Garfield, & Millikan, 1970; Rubenstein, Lewis, & Rubenstein, 1971). These studies examined only the cases in which words and nonwords were correctly distinguished, and inferences were made from the reaction times of such recognitions as a function of certain independent variables. Briefly, the model which they have proposed suggests that the subject begins by segmenting ("quantizing") the word into letters and phonemes, and recoding the phonemes into their auditory equivalents. A first check is made on the auditory recodings (essentially pronounceability) and an item which is not pronounceable is declared to be a nonword. If the item passes this first check, the subject next considers the individual letters (orthography) for acceptable English combinations. Having passed the orthography check, the subject pays attention to lexical meaning. If the item has meaning, it is accepted as a word. Evidence is also presented to show that a subject's speed in responding is directly related to word familiarity or frequency, but the authors conclude on the basis of an experiment that meaning is a more important attribute of recognition than is frequency. It should be noted that these investigators actually have presented a model of temporal priorities involved in word recognition. Their model is essentially one of a series of steps at which an item either passes or fails. They do not directly consider the occurrence of an item which may be held in varying shades of doubt at each of the check points. Were such a doubtful item to occur, though we would know the temporal order of the checks, the relative importance of each of these attributes in the final decision would still be in question.

In regard to the present experiment, this model would predict that an item's (especially a nonword's) orthography and pronounceability might be related to its recognition. Among words, meaning and frequency would be expected to be important determinants of recognition.

In order to obtain some evidence relating to these various positions, the words from Experiment I were presented to an independent group of subjects to be scaled in relation to the attributes suggested above: associability, frequency, orthography, and pronounceability.

Method

All 300 words and 100 nonwords from Experiment I were used in the present experiment. These items were divided into four groups of 100 (75 words and 25 nonwords) by matching sets of four items, words and nonwords separately, and randomly assigning one from each such set to each group of 100. The criteria for matching items were mean confidence ratings from Experiment I (the primary criterion), and the discrimination and internal consistency measures mentioned in Experiments I and II. The groups of 100 words were typed in lower-case on a single page in four columns of 25, randomly assigned within a column, with a blank preceeding each word, and a rating scale at the top of the page.

Independent groups of subjects provided four types of ratings on the items. Items were rated for associability (how many other words an item brings to mind), frequency (how often the item occurs in printed English), orthographic distinctiveness (how unusual or outstanding the letters or spelling of an item are), and pronounceability (how easy an item is to pronounce). The lowest point on the rating scale represented words which were low in associability, low in frequency, low in orthographic distinctiveness, or hard to pronounce.

Instructions for the ratings were provided on a cover sheet and were all patterned after the instructions for rating orthographic distinctiveness provided by Zechmeister (1969). Subjects rated items by writing a number from 1 to 9 on the blank beside each item. Instructions asked the subjects to rate all the words. No mention was made that some of the items were not words.

Four pages (of 100 items each) were to be rated for each attribute. It was decided, however, that each subject would rate only 200 items, or two pages. Test booklets were constructed by joining pages 1 and 2 or by joining pages 3 and 4. Across subjects, the two possible orders for the two sets of pages (1-2, 2-1, 3-4, and 4-3) were alternated to balance progressive error. Each group of 200 items was scaled for each attribute by an independent group of 26 subjects. Since there were 400 items and four attributes, there was a total of 208 subjects. The subjects were drawn from the same pool as in Experiments I and II, and were tested in groups ranging in size from 1 to about 30.

As a result of clerical error, two nonwords and six words were incorrectly typed on the rating sheets. These items have not been considered in the analyses to follow. Four subjects were dropped for failure to complete their rating sheets. Four subjects were randomly selected and their data discarded to effect equally sized groups. Three subjects had clearly and consistently reversed the direction of the rating scale and their data were corrected and retained.

According to orthodox scaling procedures, data generated by the scales used in these experiments were clearly ordinal. Up to this point, however, for the sake of convenience it was assumed that the points on the scale represented true intervals as implied by the numbers 1 through 6, and statistics were used accordingly. In Experiment III where this threatened to be a more serious problem because of the lesser number of subjects, several analyses were done using medians instead of means. The consistent finding was equivalent relative results and lowered predictability, and the use of medians was discontinued.

Results

Words and nonwords produced clear differences in means on all scales. Table 2 presents the mean scaled judgment and standard deviation for the con-

Table 2

Means and Standard Deviations of the Recognition Confidence Judgments from Experiment I and Scaling Judgments from Experiment III for Nonwords and Words.

	Nonwords (98)		Words (294)	
	Mean	SD	Mean	SD
Confidence	2.92	.51	4.89	1.08
Associability	2.40	.58	4.89	1.65
Frequency	1.91	.54	4.70	2.10
Orthography	5.54	.93	4.66	1.02
Pronounceability	4.19	1.30	6.28	1.46

fidence ratings obtained on the items in Experiment I, and for each of the attribute scalings in Experiment III. The confidence ratings were done on a six-point scale, and the scalings on a nine-point scale. The results may be summarized by the statement that as compared to words, non-words were less confidently judged to be words, were less likely to remind a subject of other words, were perceived to occur less frequently in printed English, were more distinctive in orthography, and were less pronounceable.

In Table 3 are displayed the intercorrelations of all these measures for nonwords and words. With this number of cases, a correlation of .27 is statistically different from zero ($p < .01$), and all the correlations in the table pass this criterion.

As a check on how reliable any regression analyses on these data might be, the nonwords and words were ranked separately by order of mean confidence judgment from Experiment I. Nonwords and words were then split on an odd-even basis into two groups. The "odd" nonwords and words were combined, as were the "even" nonwords and words to form two groups of items, each consisting of 49 nonwords and 147 words, and these will be referred to as Group 1 and Group 2. Group 1 and Group 2 were analyzed separately by a stepwise multiple regression analysis. The mean scale values for each item on each of the four attributes were used as predictors, and the mean confidence rating given the item in Experiment I was used as the dependent variable.

Table 3

Intercorrelation of Recognition Confidence Judgments from Experiment I and Scaling Judgments from Experiment III for Nonwords and Words

		Nonwords (98)				Words (294)			
		A	F	O	P	A	F	O	P
Confidence		.59	.73	-.36	.47	.87	.86	-.53	.71
Associability	(A)		.69	-.47	.68		.93	-.62	.76
Frequency	(F)			-.48	.60			-.67	.75
Orthography	(O)				-.62				-.78
Pronounceability	(P)								

In both regressions, values from all four scales were entered as significant predictors. The resultant multiple correlations were .917 in Group 1 and .925 in Group 2. The weightings obtained in Group 1 were then used to predict confidence ratings in Group 2. Similarly, weightings obtained from the Group 2 regression were used to predict confidence ratings in Group 1, resulting in a double cross validation on independent samples of items. The predicted and actual confidence ratings correlated .915 in Group 1, and .923 in Group 2, demonstrating remarkably small shrinkage and providing evidence for the reliability of the regression analysis results.

Items were then separated into a group of 98 nonwords and another group of 294 words, and these groupings were submitted to separate stepwise multiple regression analyses. The results for these two groupings were not the same at all. For the nonwords, only scaled frequency significantly predicted Experiment I confidence ratings. The multiple correlation was, of course, the same as the simple correlation between rated confidence and scaled frequency, namely .73 (F=106.63, df=1,96, p<.01), accounting for 53% of the variance. The F to enter the next variable (associability) into the prediction equation was not significant (F=3.04, df=1,95, p>.05), and were this variable to have been entered, it would have accounted for just over 1% more of the variance.

For words, however, all four predictor variables significantly entered into the prediction of Experiment I confidence ratings. The overall multiple correlation was .89, thus accounting for 79% of the variance. Associability, frequency, and pronounceability were entered into the prediction first, second, and fourth, respectively, all with statistical significance surpassing the .01 level (Fs=906.28, 21.66, and 16.64). Orthographic distinctiveness was entered as the third variable, significant at the .05 level (*F*=4.44, *df*=1,290). The standardized beta weights for associability, frequency, orthography, and pronounceability were calculated to be .43, .42, .19, and .21 in that order. It may be concluded, then, that in the confidence ratings of word recognition, associability, frequency, orthography, and pronounceability were all significant predictors, but as seen from the order of entry into prediction, and the standardized beta weights, there is reason to believe that the first two, associability and frequency, were of somewhat more importance.

Discussion

Several variables have been suggested as important to a subject in the distinction between words and nonwords. This experiment has provided further evidence that the confidence with which a subject recognizes a stimulus to be a word may be related in some degree to all of these variables: associability, frequency, orthography, and pronounceability. Since the data were strictly correlational, no causal inference can be made. However, the data are in line with most of the theoretical conceptions discussed in the introduction, and perhaps best aligned with the position of Rubenstein, et al. (1971). That model postulates pronounceability and orthography to be the temporally most important variables. But, considering that most words would pass a check for pronounceability and orthography, these factors would not be expected to, and, in fact, did not play as important a role (though they were significant) as did associability and frequency in the recognition of words. The Rubenstein et al. model predicted :iability (insofar as this is synonomous with their construct "meaning") and frequency to be the next important factors in identifying words, in that order. In fact, associability and frequency correlated with each other, r=.93.

The big surprise in the experiment was the prediction of nonwords. While it is true that by the nature of the way in which these nonwords were constructed, all of them were pronounceable and of acceptable orthographic structure, it would still seem that if these attributes played a significant part in the recognition of words, they should surely influence the confidence of recognition of nonwords. Yet, these factors had no measureable influence in the final regression.

The associability factor added to nonword confidence of recognition at an almost significant level. It is possible that subjects were associating to the nonword as a unit. But, it might also be hypothesized that the subjects had never seen the nonword before and were, therefore, more likely associating to some portion of the nonword, as McNulty (1966) might suggest. Possibly, some of the subjects were associating to parts of some of the nonwords, and this was inconsistent both within and between subjects. If this were the case, then perhaps had all the subjects reliably associated to the same portion of the nonword or to the entire item as a unit, the associability scale would have proven to be a significant predictor of nonword recognition confidence.

The only significant predictor of nonword judgments was perceived frequency of occurrence in printed English, accounting for about half the variance. This presents some interesting questions. It is obvious, for example, that the scaled frequencies could not be accurate estimates of true frequencies, since the true frequencies for all these items were the same, namely zero. Perhaps subjects were rating the frequencies of individual letters, letter combinations, syllables, syllable combinations, or all of these. This presents an obvious empirical question, but one for which no data can be provided in this study. If the subjects really did estimate frequency on the basis of some fragments of an item, and since these were presented in a mixed list with real words, might this imply that subjects judged the frequencies of some words in the same way, i.e., by a summing of the frequency of some word parts? Another possibility is that subjects followed two strategies when judging item frequencies; judging an item as an integrated unit when it was recognized as a word (when the frequency for the unit was perceived as greater than zero), and judging it in some segmented manner when it was not recognized as a word (when the frequency of the unit was perceived to be zero). It is clear that when subjects in an experiment have been asked to judge the frequencies of words in the language, they have been assumed to be judging the frequency of the entire unit. It is just as clear that subjects could not have been reliably judging the frequencies of entire units of nonwords, all of which had frequencies of zero.

There is another interesting implication of this with regard to a matter which was raised in the Introduction to Experiment I. It was stated there that in having subjects judge low frequency words for frequency of occurrence in printed English, the implicit assumption was being made that the subjects had, in fact, seen these words before. If, however, subjects can reliably judge (by whatever means) differences in frequency for items that are not even real words, then such an implicit assumption may not be necessary after all.

References

Banks, W. P. Signal detection theory and human behavior. Psychological Bulletin, 1970, 74, 81-99.

Carroll, J. B. Measurement properties of subjective magnitude estimates of word frequency. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 722-729.

Egan, J. P. Recognition memory and the operating characteristic. (Tech. Rep. No. AFCRC-TN-58 51, AD 152650) Indiana University, Hearing and Communication Laboratory, 1958.

G. and M. Merriam Co. Webster's seventh new collegiate dictionary. Springfield, Mass.: Author, 1963.

Green, D. M. & Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966.

Lockhart, R. S. & Murdock, Jr., B. B. Memory and the theory of signal detection. Psychological Bulletin, 1970, 74, 100-109.

McNicol, D. & Ryder, L. A. Sensitivity and response bias effects in the learning of familiar and unfamiliar associations by rote or with a mnemonic. Journal of Experimental Psychology, 1971, 90, 81-89.

McNulty, J. A. A partial learning model of recognition memory. Canadian Journal of Psychology, 1966, 20, 302-315.

Rubenstein, H., Garfield, L., & Millikan, J. . Homographic entries in the internal lexicon. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 487-494.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 645-657.

Smith, E. E. & Haviland, S. E. Why words are perceived more accurately than nonwords: Inference versus unitization. Journal of Experimental Psychology, 1972, 92, 59-65.

Stein, J. & Urdang, L. (Eds.) The Random House dictionary of the English language. New York: Random House, 1967.

Underwood, B. J. Word recognition memory and frequency information. Journal of Experimental Psychology, 1972, in press.

Underwood, B. J. & Freund, J. S. Errors in recognition learning and retention. Journal of Experimental Psychology, 1968, 78, 55-63.

Underwood, B. J. & Freund, J. S. Word frequency and short-term recognition memory. American Journal of Psychology, 1970, 83, 343-351.

Zechmeister, E. B. Orthographic distinctiveness. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 754-761.

Appendix

These are the 400 stimuli which were used in the experiments. First are the 100 nonwords in alphabetical order. After these are the 300 words in alphabetical order. The following information is listed with each item: the mean Experiment I recognition confidence rating (Mn), the standard deviation of the ratings (SD), an internal consistency measure (R) as described in the Results section of Experiment I, and, for those items which were selected for the subtest (Experiment II), a number indicating ordinal position on the test (P).

Nonwords

Item	Mn	SD	R	P	Item	Mn	SD	R	P
acromolal	3.12	1.26	.40	--	comectial	2.86	1.10	.63	--
agremagous	3.46	1.26	.62	--	corkny	2.80	1.36	.58	--
antiplas	2.92	1.21	.63	--	coure	3.36	1.31	.55	93
aochrome	3.44	1.27	.59	--	cumalink	2.50	1.02	.63	--
arrate	3.93	1.39	.55	--	defamable	4.78	1.36	.23	--
autostism	3.72	1.27	.52	35	depgeny	2.46	1.05	.64	--
baratia	3.26	1.19	.67	--	dicretule	2.76	1.12	.65	--
beweed	2.90	1.22	.56	--	disler	2.62	1.01	.58	--
bipaster	3.01	1.26	.57	5	dropant	2.56	1.12	.61	24
botsony	2.63	1.13	.66	--	ebarniter	2.68	1.08	.70	--
britching	3.64	1.39	.57	--	eltuless	2.58	1.06	.70	--
canureflow	2.68	1.04	.66	--	enblear	2.85	1.12	.59	96
ceiloplaty	3.12	1.17	.53	--	falfold	2.94	1.08	.66	--
chiless	2.76	1.30	.56	68	faperemia	2.72	1.10	.64	--
clinomible	2.52	1.05	.71	--	fixchen	2.36	1.08	.57	--
[illegible]aesly	2.42	1.04	.64	--	foolersion	2.29	0.99	.59	11

Item	Mn	SD	R	P	Item	Mn	SD	R	P
fusigenic	3.73	1.19	.36	--	outmitful	2.60	1.24	.58	62
gatebrand	3.24	1.34	.67	--	pased	2.86	1.48	.52	--
glufe	2.20	1.03	.60	--	paubub	2.28	1.00	.66	--
guarddeb	1.97	0.96	.63	65	persavort	3.08	1.21	.62	54
handman	4.32	1.43	.44	19	pigful	3.23	1.51	.44	15
hinsie	2.32	1.06	.70	--	plabage	2.74	1.15	.67	--
hould	2.88	1.16	.52	--	potomite	3.34	1.18	.59	--
hyplexion	3.44	1.36	.59	28	prieng	2.42	1.08	.73	--
imes	2.48	1.13	.56	--	prizeling	3.49	1.44	.36	--
indrabund	2.96	1.20	.61	--	puncfight	2.44	0.94	.64	--
ineffity	3.96	1.37	.54	74	radiofacy	2.76	1.12	.68	--
jigbill	3.12	1.34	.52	--	reilf	2.10	0.99	.43	--
knavagle	2.60	1.18	.64	--	relatize	3.13	1.48	.49	--
kuhead	2.44	1.09	.68	--	reweat	2.74	1.15	.58	--
leucoin	3.14	1.21	.60	51	rochead	2.88	1.26	.67	--
lightpose	3.50	1.28	.43	--	sabcwtra	2.30	1.12	.61	--
lordgly	2.22	1.05	.58	33	salcmelor	2.50	1.10	.70	--
mamazedite	2.79	1.22	.57	--	seblement	2.88	1.25	.56	--
meild	2.60	1.12	.65	--	seeve	3.80	1.75	.51	--
metenetion	2.93	1.10	.60	--	sheal	3.54	1.41	.63	--
misno	2.50	1.08	.65	--	shured	3.10	1.27	.52	86
moul	3.38	1.25	.58	--	sinmersion	2.76	1.19	.63	--
neotatin	3.16	1.25	.65	--	snaptor	2.75	1.07	.62	89
nonquasity	3.51	1.25	.55	--	sparkhouse	3.96	1.43	.48	--
ordiful	2.85	1.13	.67	63	spoter	3.80	1.81	.40	52

Item	Mn	SD	R	P	Item	Mn	SD	R	P
stedkoon	2.13	0.97	.69	16	tumcier	2.82	1.13	.71	--
stram	3.38	1.30	.58	67	unmanal	3.00	1.25	.56	73
subting	2.82	1.19	.59	--	untercron	2.66	1.22	.67	--
syroly	2.84	1.09	.58	2	vatadown	2.50	1.01	.59	--
tancannose	2.65	1.09	.66	--	virclosy	2.84	1.16	.64	--
terer	2.54	1.06	.69	75	wailjoin	2.64	1.14	.60	--
thimmery	2.96	1.23	.61	--	whitper	2.47	1.05	.64	--
toastsect	2.65	1.05	.63	--	wiltial	2.74	1.08	.64	27
tradured	3.14	1.27	.62	--	yachtfast	2.99	1.28	.46	--

Words

Item	Mn	SD	R	P	Item	Mn	SD	R	P
abdomen	5.98	0.12	.16	--	asinine	5.54	1.16	.17	--
absolution	5.76	0.82	.25	--	assuming	5.97	0.24	.05	--
accipiter	3.51	1.19	.44	--	atlas	5.99	.07	.12	--
adorable	5.99	0.14	.08	--	audile	3.96	1.35	.39	--
aerobic	5.65	0.83	.27	--	bacterium	5.62	0.94	.17	--
afterbrain	3.32	1.55	.38	--	ballroom	5.99	0.14	.02	--
albatross	5.90	0.57	.11	--	barrette	5.28	1.17	.16	--
aloft	5.86	0.59	.20	--	bateau	4.40	1.52	.37	7
ammonite	4.60	1.22	.52	--	benevolent	5.97	0.36	.10	--
anaclitic	3.74	1.36	.48	26	bezant	3.02	1.23	.43	60
anticipant	4.70	1.58	.31	--	bilabiate	3.24	1.34	.40	--
apeak	3.43	1.41	.42	--	blintze	4.26	1.82	.30	14
appendix	6.00	0.00	-	--	bluebird	5.92	0.48	.14	--
arachnid	4.86	1.50	.31	29	bogie	5.14	1.27	.35	--

Item	Mn	SD	R	P	Item	Mn	SD	R	P
bookseller	5.51	1.15	.20	--	coolie	5.52	0.99	.24	--
breathe	5.92	0.53	.11	--	corrody	3.24	1.24	.48	--
bruiser	5.36	1.09	.32	--	crasis	2.80	1.18	.53	--
bugleweed	4.20	1.45	.49	55	crewel	4.24	1.70	.34	12
bureaucrat	5.92	0.55	.19	--	crossbill	4.42	1.35	.49	71
caird	3.56	1.27	.54	--	cruzeiro	3.21	1.33	.43	69
calla	3.04	1.27	.50	36	czar	6.00	0.00	-	--
canaille	3.56	1.32	.47	22	dapper	5.75	0.63	.32	--
captivity	6.00	0.00	-	--	deaf	5.98	0.21	-.09	--
castanet	5.44	1.14	.21	50	deciliter	4.24	1.42	.43	82
cater	5.86	0.53	.11	--	denotation	5.87	0.57	.13	--
centrum	4.48	1.24	.41	--	despise	5.97	0.20	.17	--
chalaza	2.88	1.11	.54	92	deuterium	5.24	1.07	.22	--
cherish	5.99	0.07	.12	--	dime	6.00	0.00	-	--
choosy	5.61	0.94	.27	--	directed	6.00	0.00	-	--
chrysalid	3.93	1.46	.44	81	disconcert	5.69	0.86	.24	--
circumcise	5.97	0.20	.15	--	disinherit	5.92	0.41	.18	--
cockneyfy	3.20	1.55	.43	100	dogmatic	5.99	0.07	.11	--
cole	4.44	1.48	.28	--	dorm	5.40	1.35	.17	--
coloratura	3.90	1.50	.45	57	downgrade	5.84	0.58	.19	--
compressed	5.99	0.07	-.01	--	drawplate	4.08	1.45	.46	23
concurrent	5.96	0.24	.13	--	eager	5.99	0.14	-.03	--
confluence	4.66	1.29	.34	10	ecdysiast	3.23	1.41	.44	77
connote	5.42	1.09	.20	66	effector	5.18	1.16	.26	--
contrasty	2.74	1.60	.39	--	elemental	5.74	0.83	.20	--

Item	Mn	SD	R	P
enfold	5.42	1.07	.17	--
enteron	3.24	1.32	.35	--
epicurean	5.73	0.79	.21	--
etiolate	3.74	1.22	.44	88
eventuate	4.45	1.43	.44	4
exciting	6.00	0.00	-	--
exocarp	3.16	1.29	.46	--
explode	6.00	0.00	-	--
facula	3.58	1.23	.48	21
farthing	5.68	0.85	.19	--
fellah	3.46	1.85	.42	3
filefish	3.83	1.37	.46	--
firearm	5.94	0.32	.20	--
flatiron	5.24	1.27	.34	--
floatage	4.79	1.28	.42	--
forefeel	3.29	1.19	.43	--
formerly	5.92	0.61	.11	--
fourscore	5.12	1.59	.26	--
freeboard	4.30	1.47	.47	--
fully	5.96	0.41	-.09	--
gallantry	5.94	0.39	.19	--
gangue	3.44	1.34	.43	--
germinant	4.75	1.17	.40	32
gillie	3.30	1.26	.50	17
[illegible]ing	5.96	0.21	.13	--

Item	Mn	SD	R	P
goiter	5.68	0.[illegible]	.28	--
grantee	4.30	1.[illegible]	26	9
greatcoat	4.52	1.70	41	43
gristmill	4.82	1.44	32	84
gustation	3.60	1.[illegible]	[illegible]4	--
harass	5.92	0.4[illegible]	[illegible]	--
hasten	5.94	0.3[illegible]	.07	--
headgear	5.72	0.[illegible]	.22	--
hebdomad	2.53	1.11	.[illegible]	78
heroicomic	3.64	1.[illegible]	[illegible]5	76
hibachi	4.86	1.[illegible]	[illegible]4	38
hocus	5.04	1.[illegible]	.33	--
homebody	5.30	1.1[illegible]	[illegible]6	--
humid	6.00	0.00	-	--
hydraulic	5.92	0.4[illegible]	[illegible]3	--
hypotonic	4.50	1.31	36	--
imagine	5.96	0.34	.05	--
impartial	5.99	0.07	-.07	--
inanimate	5.92	0.42	.19	--
income	6.00	0.00	-	--
indigotin	3.08	1.27	.52	61
inflect	5.57	1.00	.21	--
inhumane	5.83	0.72	.15	--
insculp	3.32	1.24	[illegible]7	--
intestinal	5.98	0.17	.18	--

Item	Mn	SD	R	P	Item	Mn	SD	R	P
patty	5.60	0.96	.13	--	rabato	3.22	1.38	.40	--
penknife	5.86	0.45	.25	--	ramate	3.28	1.20	.52	--
perception	6.00	0.00	-	--	rarely	5.98	0.21	.00	--
periosteal	3.86	1.35	.37	70	redact	3.28	1.45	.41	94
petiolar	3.04	1.24	.44	90	reelect	5.71	1.00	.14	--
photometer	5.68	0.76	.31	--	regardant	3.72	1.43	.39	--
pianoforte	4.88	1.54	.33	--	reportage	4.34	1.46	.38	37
pinkeye	4.96	1.47	.37	--	reredos	2.64	1.09	.53	13
pita	3.20	1.29	.47	83	respecting	5.84	0.77	.03	--
plowable	5.28	1.18	.29	--	reverent	5.86	0.57	.22	--
pocketbook	5.96	0.24	.07	--	rigadoon	3.20	1.45	.40	--
polarity	5.92	0.44	.21	--	riser	5.66	0.82	.16	--
positivity	5.25	1.33	.13	--	rondel	3.78	1.39	.43	72
poteen	3.16	1.32	.46	--	rotatable	5.30	1.09	.32	--
practiced	6.00	0.00	-	--	runabout	5.56	1.00	.31	--
preclude	5.90	0.46	.15	44	sabadilla	2.92	1.07	.46	--
primate	5.94	0.48	.12	--	sandal	5.88	0.71	.20	--
profess	5.98	0.14	.07	--	scabby	5.40	1.06	.29	--
prolong	5.97	0.36	.10	--	scarfpin	4.42	1.62	.38	59
proposal	5.96	0.38	.14	--	schoolbag	5.56	1.01	.31	--
psychiatry	5.92	0.56	.20	--	scorify	3.25	1.31	.43	--
puffball	5.18	1.19	.45	--	screwy	5.20	1.31	.32	--
purview	3.92	1.47	.41	20	sego	3.30	1.57	.48	--
pyridoxal	3.53	1.51	.39	85	seller	5.84	0.65	.13	--
[illegible]kstep	4.74	1.49	.27	--	seminomad	4.64	1.50	.35	--

Item	Mn	SD	R	P
inventive	5.99	0.10	.09	--
irksome	5.86	0.59	.25	--
isoantigen	3.82	1.51	.40	30
javelin	5.93	0.49	.11	--
jotting	5.80	0.56	.16	--
jungle	6.00	0.00	-	--
kaross	2.69	1.25	.46	--
lady	5.98	0.25	-.02	--
landaulet	3.46	1.21	.44	--
laryngitic	4.69	1.29	.38	48
leastways	4.32	1.72	.26	56
leisurely	5.98	0.23	.14	--
licensee	5.23	1.25	.27	--
litany	5.56	0.97	.25	--
loathe	5.96	0.26	.04	--
logogriph	2.95	1.19	.38	--
loxodrome	3.74	1.15	.45	91
macrocosm	5.41	1.22	.27	--
magnitude	5.99	0.10	.16	--
maker	5.95	0.34	.08	--
mammary	5.67	0.91	.12	--
manipular	4.20	1.42	.33	--
maternity	5.99	0.07	.01	--
mealie	3.60	1.63	.36	53
[illegible]e	5.86	0.59	.14	--
mewl	3.29	1.51	.41	25
midyear	5.62	0.86	.32	--
minded	5.82	0.74	.17	--
molto	3.52	1.51	.43	45
monochord	5.20	0.95	.27	--
monzonite	3.40	1.40	.42	--
morrow	5.45	0.94	.29	--
mussiness	4.10	1.56	.53	95
mysticism	5.94	0.37	.20	--
nationwide	5.80	0.77	.19	--
necessity	5.97	0.36	-.02	--
nitrous	5.47	1.07	.30	6
norther	2.83	1.57	.34	98
nuance	5.38	1.15	.30	41
numismatic	4.60	1.41	.32	79
oddball	5.64	0.87	.24	--
ogham	2.58	1.06	.45	--
opener	5.94	0.33	.15	--
optimize	5.66	0.77	.22	--
otherguess	2.32	1.30	.38	99
overproof	4.15	1.57	.35	--
pabulum	4.78	1.46	.22	--
paralogism	4.35	1.27	.39	64
parka	5.79	0.72	.13	--
partridge	5.95	0.33	.07	--

Item	Mn	SD	R	P
serviceman	5.94	0.29	.14	--
sever	5.83	0.63	.17	--
shakily	5.72	0.85	.27	--
shiftily	5.58	0.93	.18	--
sidewise	4.72	1.65	.23	--
silkweed	5.34	1.09	.35	--
sitar	5.57	1.05	.27	--
sleepless	5.94	0.30	.14	--
sloppily	5.76	0.61	.31	--
smashing	5.94	0.39	.07	--
snuffbox	5.70	0.83	.17	--
somalo	2.86	1.25	.52	46
sorgo	2.87	1.17	.38	--
southernly	4.54	1.94	.06	--
speech	5.98	0.21	-.07	--
splore	2.84	1.20	.38	--
squadron	5.97	0.24	.17	--
stableness	5.06	1.74	.15	--
stamper	5.22	1.13	.31	--
steric	4.30	1.29	.36	--
stingaree	2.78	1.35	.33	--
stomachic	3.04	1.45	.44	--
strophe	4.60	1.50	.40	--
stylebook	5.00	1.28	.34	--
s[illegible]atum	5.48	0.93	.18	--

Item	Mn	SD	R	P
suffuse	5.14	1.24	.24	--
supertax	3.74	1.57	.43	39
surpassing	5.98	0.19	.10	--
swam	5.83	0.76	.03	--
swellfish	3.92	1.56	.44	--
tackiness	5.28	1.09	.21	--
takeoff	5.66	1.07	.21	--
tartaric	4.62	1.34	.40	80
teahouse	5.62	0.80	.30	--
tensity	4.86	1.49	.24	--
thalamic	4.36	1.35	.44	97
theorist	5.96	0.23	.15	--
thurl	2.99	1.24	.40	--
tighten	5.99	0.07	.06	--
tinting	5.87	0.44	.21	--
tonal	5.40	1.13	.28	8
tourniquet	5.80	0.68	.16	--
transcript	6.00	0.00	-	--
transshape	3.38	1.42	.48	--
treetop	5.65	0.93	.24	--
tripoli	4.78	1.51	.36	--
tropopause	3.56	1.49	.50	40
trustful	5.76	0.90	.05	--
turnabout	5.68	0.72	.25	--
ubiquitous	5.44	1.18	.20	1

Item	Mn	SD	R	P	Item	Mn	SD	R	P
unction	4.87	1.60	.32	87	warren	4.42	1.61	.34	31
undies	3.76	1.81	.43	34	waterfowl	5.54	1.03	.31	--
unilinear	5.50	1.02	.29	--	waybill	4.04	1.40	.37	49
urchin	5.98	0.14	.20	--	whangee	2.72	1.19	.43	--
valueless	5.67	0.88	.18	--	whicker	4.38	1.80	.19	--
vasculum	4.78	1.09	.33	--	whoso	2.82	1.76	.35	42
verseman	3.96	1.33	.50	--	wivern	2.91	1.14	.54	--
vexillate	3.86	1.26	.44	--	woozily	4.78	1.44	.40	58
vilifier	4.08	1.52	.34	18	worrywart	4.88	1.50	.34	--
viva	5.21	1.34	.28	--	writer	6.00	0.00	-	--
wanderlust	4.96	1.44	.31	47	zipper	6.00	0.00	-	--

DISTRIBUTION LIST

NAVY

4 Dr. Marshall J. Farr
Director, Personnel and Training
Research Programs
Office of Naval Research
Arlington, VA 22217

1 Director
ONR Branch Office
495 Summer Street
Boston, MA 02210

1 Director
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Director
ONR Branch Office
536 South Clark Street
Chicago, IL 60605
ATTN: Dr. M. Bertin

6 Director
Naval Research Laboratory
Code 2627
Washington, DC 20390

12 Defense Documentation Center
Cameron Station, Building 5
5010 Duke Street
Alexandria, VA 22314

1 Chairman
Behavioral Science Department
Naval Command and Management Division
U.S. Naval Academy
Luce Hall
Annapolis, MD 21402

1 Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054
ATTN: Dr. G. D. Mayo

1 Chief of Naval Training
Naval Air Station
Pensacola, FL 32508
ATTN: CAPT Allen E. McMichael

1 Chief
Bureau of Medicine and Surgery
Code 513
Washington, DC 20390

1 Commander Naval Air Reserve
Naval Air Station
Glenview, IL 60026

1 Commander
Naval Air Systems Command
Navy Department, AIR-413C
Washington, DC 20360

1 Commanding Office
Naval Personnel and Training
Research Laboratory
San Diego, CA 92152

1 Head, Personnel Measurement Staff
Capital Area Personnel Service Office
Ballston Tower #2, Room 1204
801 N. Randolph Street
Arlington, VA 22203

1 Program Coordinator
Bureau of Medicine and Surgery (Code 71G)
Department of the Navy
Washington, DC 20390

1 Research Director, Code 06
Research and Evaluation Department
U.S. Naval Examining Center
Building 2711 - Green Bay Area
Great Lakes, IL 60088
ATTN: C.S. Winiewicz

1 Superintendent
Naval Postgraduate School
Monterey, CA 93940
ATTN: Library (Code 2124)

1 Technical Director
Naval Personnel Research and
Development Laboratory
Washington Navy Yard
Building 200
Washington, DC 20390

Technical Director
Personnel Research Division
Bureau of Naval Personnel
Washington, DC 20370

Technical Library (Pers-11B)
Bureau of Naval Personnel
Department of the Navy
Washington, DC 20360

Technical Library
Naval Ship Systems Command
National Center
Building 3 Room 3
S-08
Washington, DC 20360

Technical Reference Library
Naval Medical Research Institute
National Naval Medical Center
Bethesda, MD 20014

Behavioral Sciences Department
Naval Medical Research Institute
National Naval Medical Center
Bethesda, MD 20014

COL George Caridakis
Director, Office of Manpower Utilization
Headquarters, Marine Corps (A01H)
MCB
Quantico, VA 22134

Mr. George N. Graine
Naval Ship Systems Command
(SHIPS 03H)
Department of the Navy
Washington, DC 20360

CDR Richard L. Martin, USN
COMFAIRMIRAMAR F-14
NAS Miramar, CA 92145

Mr. Lee Miller (AIR 413E)
Naval Air Systems Command
5600 Columbia Pike
Falls Church, VA 22042

Dr. James J. Regan
Code 55
Naval Training Device Center
Orlando, FL 32813

1 Dr. A. L. Slafkosky
Scientific Advisor (Code Ax)
Commandant of the Marine Corps
Washington, D.C. 20380

1 LCDR Charles J. Theisen, Jr., MSC, USN
CSOT
Naval Air Development Center
Warminster, PA 18974

1 Dr. Harold Booher
NAVAIR 415C
6500 Columbia Pike
Falls Church, VA 22042

ARMY

1. Behavioral Sciences Division
Office of Chief of Research and Development
Department of the Army
Washington, DC 20310

1 U.S. Army Behavior and Systems Research Laboratory
Rosslyn Commonwealth Building, Room 239
1300 Wilson Boulevard
Arlington, VA 22209

1 Director of Research
U.S. Army Armor Human Research Unit
ATTN: Library
Building 2422 Morade Street
Fort Knox, KY 40121

1 Director
Behavioral Sciences Laboratory
U.S. Army Research Institute of Environmental Medicine
Natick, MA 01760

1 U.S. Army Research Institute
Room 239
Commonwealth Building
1300 Wilson Boulevard
Arlington, VA 22209
ATTN: Dr. R. Dusek

AIR FORCE

1 AFHRL (TR/Dr. G. A. Eckstrand)
Wright-Patterson Air Force Base
Ohio 45433

1 AFHRL (TRT/Dr. Ross L. Morgan)
Wright-Patterson Air Force Base
Ohio

1 AFHRL/MD
701 Prince Street
Room 200
Alexandria, VA 22314

1 AFOSR (NL)
1400 Wilson Boulevard
Arlington, VA 22209

1 COMMANDANT
USAF School of Aerospace Medicine
ATTN: Aeromedical Library (SCL-4)
Brooks AFB, TX 78235

1 Personnel Research Division
AFHRL
Lackland Air Force Base
San Antonio, TX 78236

1 Headquarters, U.S. Air Force
Chief, Personnel Research and Analysis Division (AF/DPXY)
Washington, DC 20330

1 Research and Analysis Division
AF/DPXYR Room 4C200
Washington, DC 20330

1 Headquarters Electronic Systems Division
ATTN: Dr. Sylvia R. Mayer/MCIT
LG Hanscom Field
Bedford, MA 01730

DOD

1 Mr. Joseph J. Cowan, Chief
Psychological Research Branch (P-1)
U.S. Coast Guard Headquarters
400 Seventh Street, SW
Washington, DC 20590

1 Dr. Charles Ullman
Chief of Counseling, Training Programs
OSD (M&RA)
The Pentagon, Room 2C252
Washington, DC 20301

OTHER GOVERNMENT

1 Dr. Alven E. Goins, Chief
Personality and Cognition Research Section
Behavioral Sciences Research Branch
National Institute of Mental Health
5600 Fishers Lane
Rockville, MD 20852

1 Dr. Andrew R. Molnar
Computer Innovation in Education Section
Office of Computing Activities
National Science Foundation
Washington, DC 20550

MISCELLANEOUS

1 Dr. Scarvia Anderson
Executive Directory for Special Development
Educational Testing Service
Princeton, NJ 08540

1 Professor John Annett
The Open University
Waltonteale, BLETCHLEY
Bucks, ENGLAND

1 Dr. Richard C. Atkinson
Department of Psychology
Stanford University
Stanford, CA 94305

1 Professor Mats Bjorkman
University of Umea
Department of Psychology
Radhuseplanaden 2
S-902 47 UMEA/SWEDEN

1 Mr. H. Dean Brown
Stanford Research Institute
333 Ravenswood Avenue
Menlo Park, CA 94025

1 Dr. A. Collins
Bolt Beranek and Newman
50 Moulton Street
Cambridge, MA 02138

1 Dr. Kenneth E. Clark
University of Rochester
College of Arts and Sciences
River Campus Station
Rochester, NY 14627

1 ERIC
Processing and Reference Facility
4833 Rugby Avenue
Bethesda, MD 20014

1 Dr. Victor Fields
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. Robert Glaser
Learning Research & Development Center
University of Pittsburgh
Pittsburgh, PA 15213

1 Dr. Albert S. Glickman
American Institutes for Research
8555 Sixteenth Street
Silver Spring, MD 20910

1 Dr. Duncan N. Hansen
Center for Computer-Assisted Instruction
Florida State University
Tallahassee, FL 32306

1 Human Resources Research Organization
Division #3
Post Office Box 5787
Presidio of Monterey, CA 93940

1 Human Resources Research Organization
Division #4, Infantry
Post Office Box 2086
Fort Benning, GA 31905

1 Human Resources Research Organization
Division #5, Air Defense
Post Office Box 6057
Fort Bliss, TX 79916

1 Library
HumRRO Division Number 6
P.O. Box 428
Fort Rucker, AL 36360

1 Dr. Lawrence B. Johnson
Lawrence Johnson and Associates, Inc.
2001 "S" Street, NW
Suite 502
Washington, DC 20009

1 Dr. Robert R. Mackie
Human Factors Research, Inc.
Santa Barbara Research Park
6780 Cortona Drive
Goleta, CA 93017

1 Mr. Luigi Petrullo
2431 North Edgewood Street
Arlington, VA 22207

1 Dr. Robert D. Pritchard
Assistant Professor of Psychology
Purdue University
Lafayette, IN 47907

1 Dr. Diane M. Ramsey-Klee
R-K Research & System Design
3947 Ridgemont Drive
Malibu, CA 90265

1 Dr. Joseph W. Rigney
Behavioral Technology Laboratories
University of Southern California
3717 South Grand
Los Angeles, CA 90007

1 Dr. Leonard L. Rosenbaum, Chairman
Department of Psychology
Montgomery College
Rockville, MD 20850

1 Dr. George E. Rowland
Rowland and Company, Inc.
Post Office Box 61
Haddonfield, NJ 08033

1 Dr. Benjamin Schneider
Department of Psychology
University of Maryland
College Park, MD 20742

1 Dr. Robert J. Seidel
Human Resources Research Organization
300 N. Washington Street
Alexandria, VA 22314

1 Dr. Arthur I. Siegel
Applied Psychological Services
Science Center
404 East Lancaster Avenue
Wayne, PA 19087

1 Dr. David Weiss
University of Minnesota
Department of Psychology
Elliott Hall
Minneapolis, MN 55455

1 Mr. Dennis J. Sullivan
Training Department
Hughes Aircraft Co.
P.O. Box 90515
Los Angeles, CA 90009

1 Dr. Frank J. Harris
Chief, Training Division
U.S. Army Research Institute
1300 Wilson Boulevard
Arlington, VA 22209

1 LCOL Austin W. Kibler, Director
Human Resources Research Office
ARPA
1400 Wilson Boulevard
Arlington, VA 22209

1 Century Research Corporation
4113 Lee Highway
Arlington, VA 22207